Translate a business problem into an AI and data science solution

To translate a business problem into an AI and data science solution, you need to understand the problem, the data analysis goals and metrics, and the mapping to one or more business patterns. Most importantly, you need to understand what the business expects to gain from the data analysis and how the results of the analysis will be used.

When a data scientist interviews a line-of-business expert about a business problem, the data scientist listens for key words and phrases. The data scientist breaks the problem into a process flow that always includes an understanding of the business problem, an understanding the data that is required, and the types of artificial intelligence (AI) and data science techniques that can solve the problem. Together, this information drives an iterative set of thought experiments, modeling techniques, and evaluation against the business goals.

The focus must stay on the business. Bringing in technology too soon can guide the solution to a technology, and the actual business problem might be forgotten or not fully answered.

AI and data science require a level of precision that is important to capture up front:

- Describe the problem to be solved
- Specify all the business questions as precisely as possible

- Determine any other business requirements, such as not losing a customer while increasing cross-sell opportunities
- Specify the expected benefits in business terms, such as reducing churn among high-value customers by 10%

Determine data analysis goals

After the business goals are clear, you need to translate them into data analysis goals and activities. For example, if the business objective is to reduce churn, you might set these analysis goals:

- Identify high-value customers based on recent purchase data.
- Build a model by using available customer data to predict the likelihood of churn for each customer.
- Rank customer based on churn propensity and customer value.

A key question to answer in follow-on activities is whether the data from the customer contains the correct information to answer the business problem. It's also important to consider how you might act on the results of this analysis to support the business goals. How do you consume and deploy the results of the analytics, and what action do you take within the business?

The business problem often requires post-processing the outputs of the AI and data science models. For example, the list of customers who might churn needs to be ordered to determine a next-best action. A low lifetime-value customer with a high churn might not be worth spending time and money to retain. Customers with high-lifetime value might take priority over lower-value customers with higher churn scores.

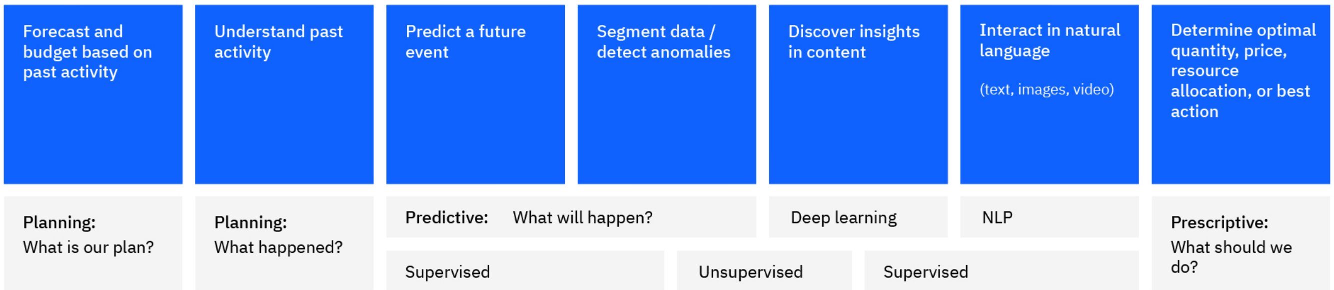
Data analysis success criteria

To keep the data analysis on track, define success in technical terms:

- Describe the methods for model assessment, such as accuracy and performance.
- Define benchmarks for evaluating success, being sure to provide specific numbers.
- Define subjective measurements as best you can and determine the arbiter of success.

Patterns of analytic business problems

At a broad level, solving business problems with data and AI requires a combination of business analytic patterns. Although they're often indicated in a lifecycle, the actual application of these patterns is more iterative in support of the overall business objective:



Click to expand the image

When you forecast and budget based on past activity, you answer the question "What is our plan?" This stage is forward thinking and often focuses on budgeting and related milestones to create more accurate plans, budgets, and forecasts that are based on past activity.

To understand past activity, you provide descriptive answers. Answer the question, “What happened?” Examples include the exploration of financial results to find root causes, directed “what-if” analysis, year-over-year, results-vs-plan, and forecasting by using numeric input and time series.

Next, you predict what actions might happen in a sequence or what events happen at the same time. Examples include churn, financial forecasts, next best action, and predicting a device failure. Predicting a future event is a supervised data science activity, as it requires a baseline of historical data to affect outcomes.

As you segment data and detect anomalies, you predict a future event based on related data elements. This exercise is unsupervised, as the patterns that are exposed in support of the business objective might not have been known before the exercise. Examples include creating market segments, detecting fraud, and identifying entities in images.

To discover insights in content, you use content analytics, deep learning, and other techniques for text, images (faces), audio (voice printing), and video and other unstructured content. You can use techniques such as feature extraction, sentiment analysis, and deep learning to extract meaning from unstructured content in support of the business objectives.

While the ability to interact in a natural language is a possible target deployment of AI and data science models, the actual interaction is another focus area of AI and data science development. Natural language processing (NLP) analyzes unstructured text data. Many prebuilt NLP models allow the analysis of sentiment, emotion, attitude, and personality. Some prebuilt models can also consume ontologies and taxonomies to extract conceptual terms and entities from text. Often, custom NLP models are required because the business use case requires specific concepts to be modeled, and the related data is often written in a domain- or industry-specific vocabulary.

To determine optimal quantity, price, resource allocation, or best action, you use decision optimization techniques that use forecasts of future scenarios. An example in next-best-action and call-center applications is ordering lists by business priority with weights from predictive analytics.

Start with an example

A bank wants to optimize its loan underwriting procedures. Currently, it applies filters on loan applications that automatically reject the riskier ones. However, the bank is still approving too many applications that run into repayment issues.

The bank collects a large amount of data for each approved loan: 146 fields. These fields can be split into a few distinct groups:

- Loan demographics, such as the amount, the term, the interest rate, and the reason for loan.
- Applicant demographics, such as age, salary, employment length, and home ownership.
- Numerous risk factors, such as the number of public records, credit card delinquency, and bankruptcy. Roughly 70 sparsely populated risk fields cover these risk factors.

The goal is to create a predictive model to identify loans that might be bad loans. However, in the raw data, no one field indicates whether the loan is good or bad.

Rachel, a data scientist, sits down with the customer to understand the customer's business problem. Her first goal is to articulate the overall business problem: “Can I detect attributes about a person or a type of loan that can be used to flag a risky loan that needs to be processed by my loan underwriting team?”

Rachel can then expand the business problem to the next level of detail: what needs to be done to answer the problem. She needs to be able to do these tasks:

- Detect what a risky loan is or detect which loans ran into problems
- Segment loans into different categories by using information such as the purpose for the loan, the loan amount, and the term of the loan
- Sort customers into different groups based on their demographic data
- Discover patterns that cross all three types

Understanding the business problem is fundamental to AI and data science solutions. To break complex needs into manageable and repeatable methods to solve the problem, you need a clear understanding of the problem, the metrics to baseline and validate, and the patterns to solve the problem.

What's next

- Prepare data for AI and data science (/garage/method/practices/code/data-preparation-ai-data-science)
- Select and develop an AI and data science model (/garage/method/practices/reason/model-selection-development-ai-data-science)

How helpful was this article:

★★★★★

Add comments ...

Authored by

Tommy Eunice

IBM, Data Scientist, IBM Data and Artificial Intelligence

Edd Biddle

IBM, Watson Senior Data Scientist

Paul Christensen

IBM, Lead Architect, IBM Analytics Services, OpenGroup Distinguished IT Specialist